

Appendix 2: Description of modelling methodology used in BICCO-Net II

1. Analysis of national index and species abundance data

1.1 Abundance data sources

We extracted records of abundance for species of butterflies, moths, aphids, birds, and mammals from the data bases of various long-term monitoring schemes within Britain (Table 1). For each species, these data comprised estimates of abundance for combinations of sites and years. For the birds, moths, mammals and aphids, species data were modelled to create annual population indices, accounting for missing counts (Freeman & Newson, 2008), giving a total of 498 national abundance indices which we then proceeded to analyse on a logarithmic scale.

The spread of sites at which data were available for other species groups was not considered suitable to form national indices. For each of these species, within-year data from any relevant site were combined to give a single value for each year: these annual values were then analysed on a site by site basis.

1.2 Weather data

We have examined the impact of two key weather variables, temperature and precipitation, on species abundance data. We derived national averages of monthly temperature (mean, minimum and maximum) and precipitation from the Met Office UKCP09 5 km x 5 km gridded datasets (<http://www.metoffice.gov.uk/climate/uk/datasets/> - Perry & Hollis, 2005) from all calendar years from 1910 to 2012. A comparison of mean, minimum and maximum temperature values indicated that these were highly correlated, and in discussion with the Steering Group it was agreed we should not include either minimum or maximum temperature in our analyses, restricting these analyses to mean temperature and precipitation.

For those species whose abundance was analysed at an individual site level, the most appropriate site-specific rainfall and temperature data were used.

For each species, it was then necessary to align the weather data to ensure analysis was based only on the months that could have influenced the response variable. This was done by identifying either the time period of the survey(s) or, for moths and aphids, the flight period of adults of a particular species within a broader survey window (Fig 1). This process allowed us to construct two sequences (one each for mean temperature and rainfall) of monthly weather variables for each species, starting from the last weather record that could reasonably have affected the recorded response variable and working backwards in time.

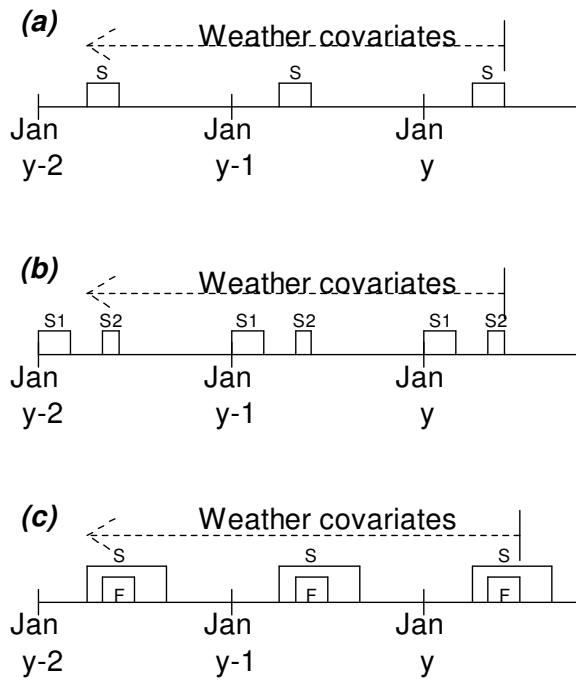


Figure 1 Illustration of how weather covariates were selected for species with (a) a single survey period (e.g. birds and mammals except bats), (b) two survey periods (e.g. bats) or (c) species with a flight period within a more general survey period (e.g. butterflies, moths and aphids). ‘S’ indicates a survey period, ‘F’ indicates a flight period. The figure illustrates the inclusion to two years of weather covariates but this is extended for up to ten years.

Monthly weather covariates were standardised by centering on the mean value for the monthly weather variable across all years and scaled by the standard deviation before inclusion in the model.

1.3 Relating abundance to weather

Estimating the relationship between weather and species abundance data is less straightforward than might appear to be the case at first sight. The key challenge to be addressed is that the number of years for which abundance data are available is small (always less than 50) considering the number of potential explanatory variables (monthly covariates for both mean temperature and rainfall, for as many years leading up to the index record as we are willing to contend).

Many approaches have been adopted by previous authors, including by the first phase BICCO-Net team. Four broad approaches have previously been considered available:

- a) some form of stepwise selection of monthly covariates from a limited set of months;
- b) identifying a key period (for example the preceding winter) and forming a single covariate by averaging over the relevant months;
- c) using a multivariate technique to reduce the dimensionality of the explanatory variables without using information in the response variable (for example taking the first two dimensions from a principal components analysis) and including these derived variables as candidate explanatory variables;

- d) using partial least squares to simultaneously combine information in both the explanatory variables and the response variable to derive a small number of candidate explanatory variables of each species.

Each of these approaches has its own merits, but none seemed entirely satisfactory for our requirements. Approaches a) and b) lead to all except a small number of regression coefficients being set to zero, despite the fact that it would be reasonable to expect the weather in consecutive calendar months to have rather similar effects on abundance. Approaches c) and d), whilst leading to non-zero coefficients, make no explicit use of the months being consecutive, or that the effect of weather in a given month (say January) may be expected to be reduced as the number of years before the abundance data is collected increases.

The alternative approach that we have adopted is to allow the weather data from a large number of past months to be included in the model for each response variable ($y = \log(\text{abundance})$), but to constrain the regression coefficients to lie on a curve defined by a small number of parameters. The general form of the relationship we use, written to allow a maximum of ten years of weather data to be included, has the following form

$$y_t = a + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_{120}x_{120} + ct + e_t \text{ (Equation 1)}$$

in which:

a is some overall intercept;

b_i is the regression coefficient for the i^{th} weather covariate, with $i=1$ denoting the most recent and $i=120$ denoting the end of the ten year period;

c is the regression coefficient of a year of survey covariate, included to remove spurious effects due to linear trends caused by excluded effects;

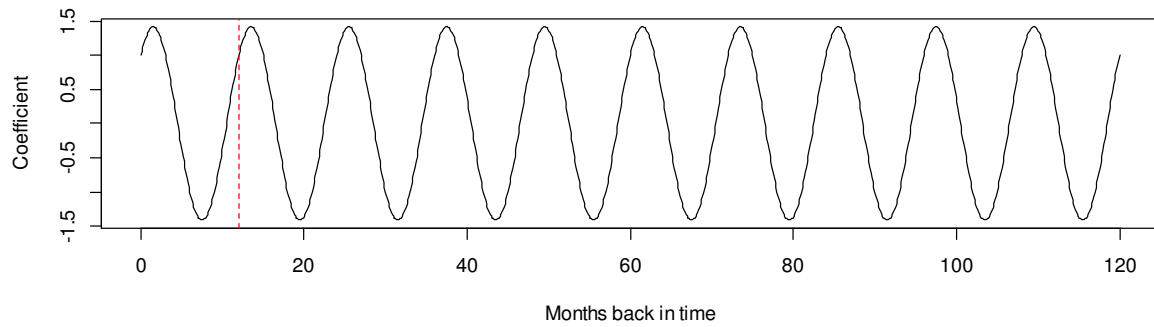
and e_t is an error term from a Normal distribution through which we allow for temporal correlation through a 1st order auto-regressive (AR1) process.

The form of the curve on which the series of regression coefficients, b_1 to b_{120} , were constrained to lie comprised a repeating, cyclic pattern combined with a decay towards zero. The cycle either repeated itself every year or every two years (Fig 2a and 2b). This allowed us to simplify the models to two parameters for each of the cyclic curves, and another parameter for the decay (Fig 2c). Mathematically, the above relationships in combination (Fig 2d) can be expressed as

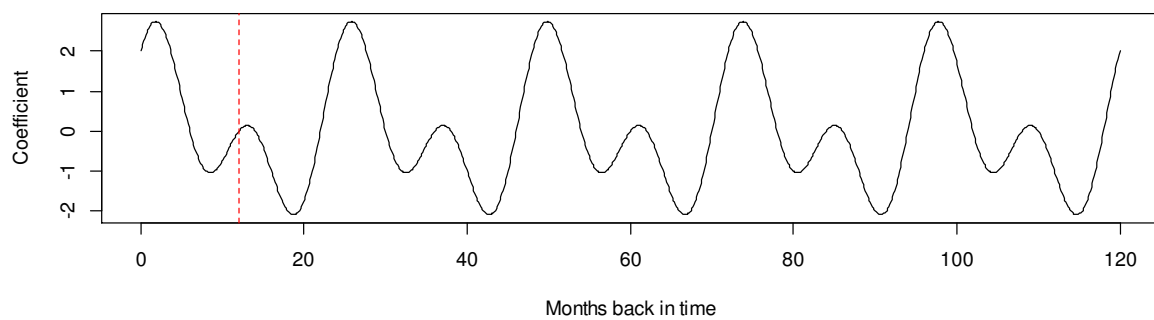
$$b_i = e^{-di} \cdot s(i),$$

where d is the decay parameter, constrained to be positive, for which large values indicate fast decay, and $s(i)$ is the cyclic term parameterised as a linear combination of sine and cosine terms of different periodicities. Note that both sine and cosine terms are required for each periodicity to allow the peak values to correspond to any month of the year.

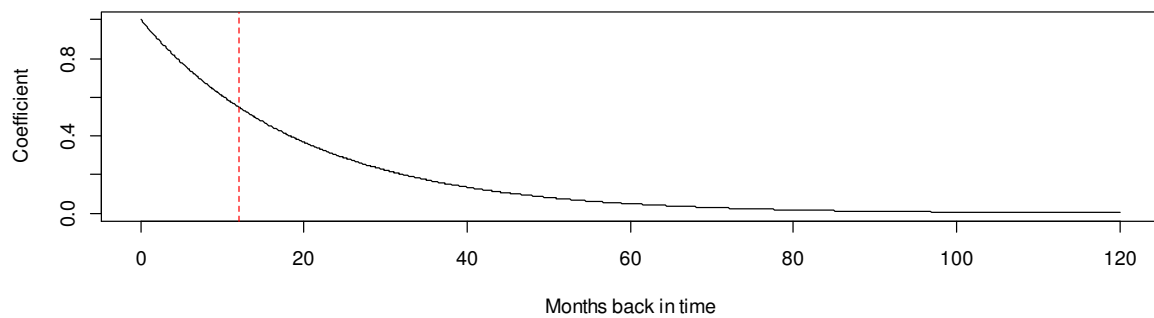
(a) Single repeating cycle per year



(b) Combination of two cycles, one repeating each year, one repeating every two years



(c) Exponential decay towards zero



(d) Product of (a) and (c), giving a repeating pattern which decays towards zero

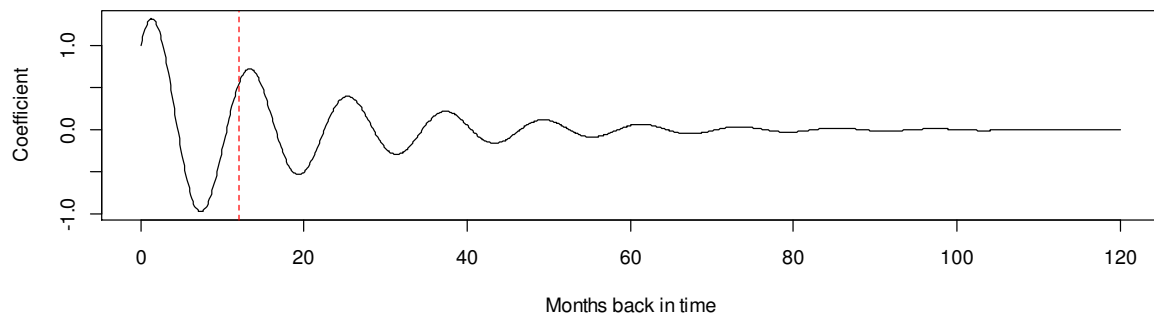


Figure 2 Plots showing how the pattern of regression coefficients for equation 1 is built up in stages. Firstly, a repeating cycle is formed from either a single term (a) or a combination of two terms (b); such cycles are then multiplied by an exponential decay (c) to produce a sequence of regression coefficients which show a repeating seasonal pattern that decays towards zero.

We compared four models of varying complexity. The first model, which we describe as the null model M_0 , assumed that population abundance did not vary with weather and hence all the regression coefficients b_i were set to zero. In this null model, the parameter d is undefined, whereas it is present in all other models. The second model, which we describe as M_2 , assumed one sine-wave cycle per year, the mathematical form for this being

$$s(i) = s_0 + s_1 \cdot \sin(2\pi i/12) + s_2 \cdot \cos(2\pi i/12) \quad (\text{Equation 2, model } M_2)$$

with parameters s_1 and s_2 of the sine and cosine terms augmented by a third parameter, s_0 , which represents the median effect of temperature throughout a seasonal cycle. This parameter allows all the regression coefficients b_i within a cycle to be positive (or negative). The third model, which we describe as M_3 , consists of two sine-wave cycles, one of period one year, and one of period six months. The net result is that the regression coefficients for M_3 have a cyclic pattern with period one year, but with greater flexibility as to the shape of the curve. The mathematical form for M_3 is

$$s(i) = s_0 + s_1 \cdot \sin(2\pi i/12) + s_2 \cdot \cos(2\pi i/12) + s_3 \cdot \sin(2\pi i/6) + s_4 \cdot \cos(2\pi i/6) \quad (\text{Equation 3, model } M_3).$$

The final model, which we describe as M_4 , consists of two sine-wave type cycles, one of period one year, one of period two years. The net result is that the regression coefficients for M_4 have a cyclic pattern with period two years, allowing for example the effect of May in the year of survey to differ from the effect of May in the previous year. The mathematical form for M_4 is

$$s(i) = s_0 + s_1 \cdot \sin(2\pi i/12) + s_2 \cdot \cos(2\pi i/12) + s_3 \cdot \sin(2\pi i/24) + s_4 \cdot \cos(2\pi i/24) \quad (\text{Equation 4, model } M_4).$$

Example results for models M_2 , M_3 and M_4 are shown in Fig 3 for one species of bird and one species of moth. The top and bottom graphs corresponding to M_2 and M_4 in the left hand column indicates that the effect of mean temperature on the national grey wagtail abundance index is primarily positive (warmer temperatures than average in any month tend to lead to higher abundances in the future), although the central figure corresponding to M_3 indicates that there is some evidence to suggest that warmer temperatures in autumn and spring lead to lower abundances in future (negative regression coefficients). The figure on the bottom right is notable for the rapid decay of the regression coefficients over time, indicating that only mean temperatures in the most recent 12 months are influencing each year's abundance index.

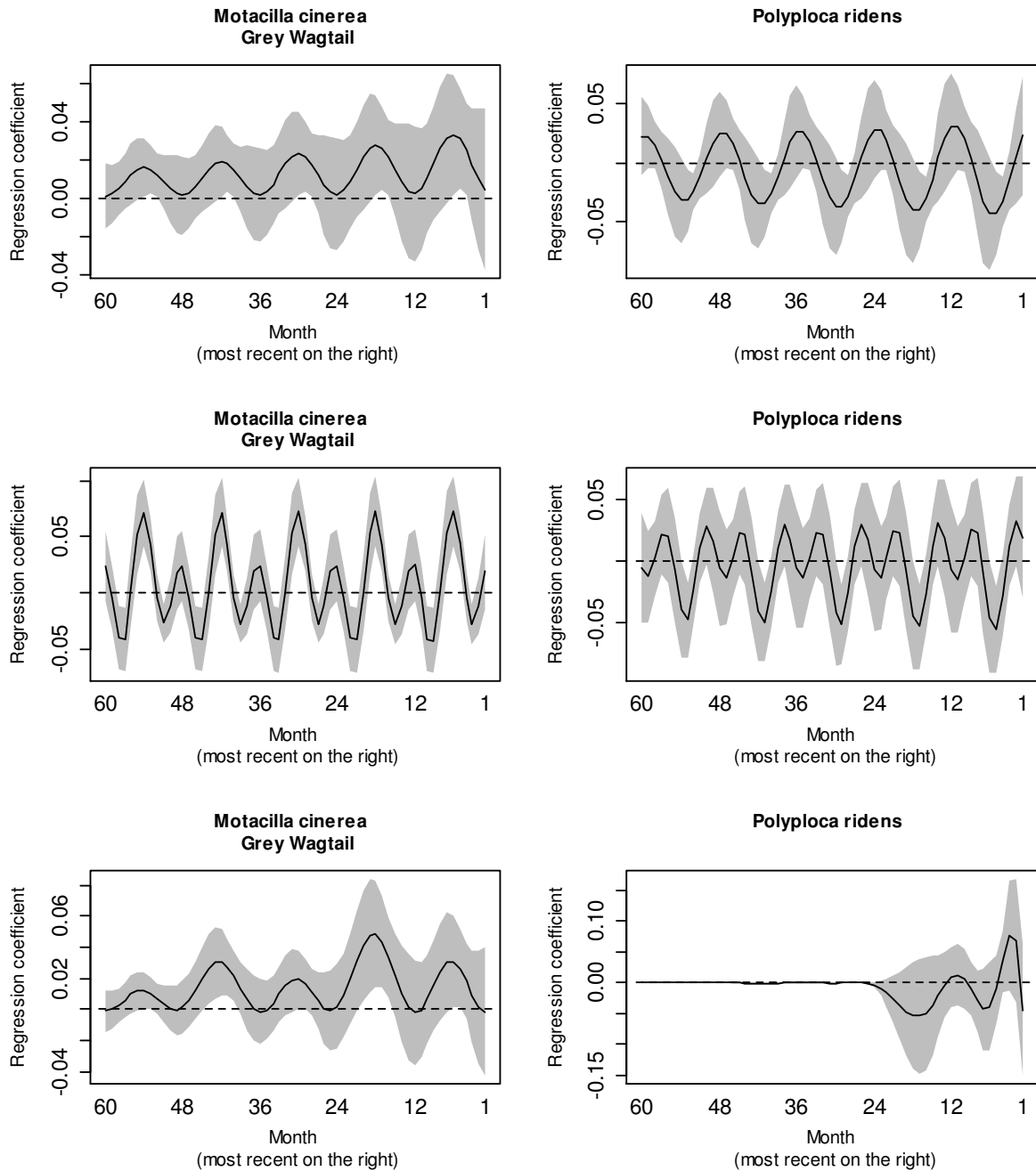


Figure 3 Regression coefficients for the mean temperature weather covariates from 60 months prior to sampling of two species (the bird grey wagtail *Motacilla cinerea* and the moth frosted green *Polyploca ridens*) from the second (top), third (middle) and fourth (bottom) types of models. Solid lines indicate the best estimates of the curves on which the regression coefficients lie; the shaded regions indicate the 95% confidence intervals.

We assessed the evidence for increasing model complexity by using chi-squared tests to examine the statistical significance of changes in log-likelihoods when adding parameters into the model (i.e. for progressing from M_0 to M_2 and from M_2 to either M_3 or M_4). Where the null model was not found to

be significantly improved upon, we demonstrate the relationship with weather using the regression coefficients from M_2 .

We also used a second method to assess the impact of weather on species abundance: this involved using multivariate techniques to summarise the weather data, then regressing the abundance indices on the summaries of the weather data. This allowed us to include temperature and precipitation into a single model, something we could not do using the technique described above because the number of parameters in the model was too high in comparison to the number of observations (one per year). We used principal components analysis (PCA) to simplify variation in monthly mean temperature and precipitation data for the two years leading up to each abundance estimate into key weather variables, the principal components. Population abundance was then modelled against these principal components.

We used generalised least squares (GLS) with a first order autoregressive correlation structure to model population growth on a logarithmic scale ($P_t = \ln(n_t / n_{t-1})$, $t \geq 2$), where n_i is the population abundance in year t against the first four PCA axes, and population abundance ($\ln(n_t)$) in the previous year

$$P_t = \alpha + b_0 \ln(n_{t-1}) + b_1 \text{PCA1}_t + b_2 \text{PCA2}_t + b_3 \text{PCA3}_t + b_4 \text{PCA4}_t + \text{error}_t \quad (\text{Equation 5}).$$

Note here that modelling population growth, $P_t = \ln(n_t / n_{t-1})$, on a logarithmic scale with the log of previous abundance as a covariate leads to the same estimates of dependence on the weather-derived principal components as if the response variable had been log abundance, since Equation 2 can be rewritten as Equation 3 by the addition of 1 to the regression coefficient of $\ln(n_{t-1})$.

$$\ln(n_t) = \alpha + (b_0 + 1) \ln(n_{t-1}) + b_1 \text{PCA1}_t + b_2 \text{PCA2}_t + b_3 \text{PCA3}_t + b_4 \text{PCA4}_t + \text{error}_t \quad (\text{Equation 6}).$$

Three separate estimates of annual population growth were produced for year t ($t \geq 2$) from this model, all starting from the first abundance estimate $\ln(n_1)$.

$$P_{Ft} = \alpha + b_0 \ln(n_{t-1}) + b_1 \text{PCA1}_t + b_2 \text{PCA2}_t + b_3 \text{PCA3}_t + b_4 \text{PCA4}_t \quad (\text{Equation 7}).$$

$$P_{Nt} = \alpha + b_0 \ln(n_{t-1}) \quad (\text{Equation 8}).$$

$$P_{Ct} = P_{Ft} - P_{Nt} \quad (\text{Equation 9}).$$

The first estimate, P_{Ft} , used the full model that included all parameters, (Equation 7). The second estimate, P_{Nt} , excluded the parameter estimates associated with the four PCA axes and indicated population growth in the absence of any influence from the weather (Equation 8). The third estimate, P_{Ct} , described the contribution that could be attributed to variation in the weather and was calculated from the difference between the first and second time series (Equation 9). This estimation of P_{Ct} (instead of estimating P_{Ct} directly from the four weather coefficients) allowed us to identify the component of population growth driven by climate taking density dependent growth into account. These growth rates were used to estimate population abundances and to assess the impact of weather over time on the species analysed.

2. Fine scale analysis of species abundance data

We previously carried out national-scale analyses of the effect of climate change on population abundances of 498 species of birds, mammals, aphids, butterflies and moths.

To complement analyses of national abundance indices of birds, mammals, aphids, butterflies and moths, we have also conducted finer-scale analyses in which the response variable contains a value for each site measured in each year rather than a single 'national' value for each year. This approach has the potential to better account for local variation in sampling and weather variables, as well as providing the opportunity to look at interactions between the effects of weather variables and some site characteristics.

Hypotheses

We presented three hypotheses to the Steering Group on the proposed focus for the fine-scale analysis as this work could be directed in more ways than we had resources to implement.

H1) Oliver et al. (2010) found that habitat heterogeneity promoted stability in the population dynamics of butterflies in Britain. They suggested that this may be due to a reduction in the impact of climate in more diverse landscapes, as heterogeneous landscape offer a wider range of microclimates that can buffer populations against climatic variation. We could further the work of Oliver et al. (2010) by increasing taxonomic coverage and directly testing the influence of habitat heterogeneity on relationship between climate and population dynamics. The associated hypothesis is that sites surrounded by more heterogeneous landscape would show a reduced impact of climate on population dynamics, as species are able to persist in these regions during times of climatic perturbation by utilising the wider variety of microhabitats available in the heterogeneous landscape.

H2) We could assess whether the amount of semi-natural habitat surrounding a site affects the relationship between climate and population dynamics. Populations in regions with low semi-natural habitat may be vulnerable to climatic variability if we assume that they are occupying a narrow part of their climatic niche. The associated hypothesis is that the impact of climate on population dynamics will be exacerbated in regions with a low proportion of semi-natural area. In addition, regions with a high amount of semi-natural areas are more likely to have immigration from nearby populations (assuming most species persist in semi-natural areas, but not in non-natural habitat), which would potentially reduce the impact of climate change on population dynamics.

H3) The impact of climate on population dynamics may vary geographically and in particular may vary based on the climate of a geographic region. Species have a specific thermal tolerance at which they can persist. Populations at the limit of their climatic niche are likely to show a stronger response to climate variability than those that are in the centre of their niche, and this may be important as many species reach their cooler thermal margin in Britain. For example, a minor drop in the temperature for a population in a site at the lower thermal tolerance limit is likely to result in a population decline, whereas the impact is likely to be minimal for a populations occupying a more central position in the climatic niche for that species. The associated hypothesis is that the relationship between climate and population dynamics will show greater prominence in cooler sites.

Following discussion with the Steering Group, we decided to concentrate resources on hypothesis H3.

Using national-scale weather models for fine-scale analysis

The national-scale national scale analyses summarises in a robust manner the cumulative effects of temperature and precipitation on a particular population or species. We used the regression coefficients from these models for each species as inputs for the fine-scale analysis to derive both a site-specific combined temperature covariate and a site-specific combined rainfall covariate for each species. For each species, this involved taking the series of regression coefficients defined by the product of the cyclic term and the decay term defined in Section 1.3 above, b_j , multiplying each coefficient by each site-specific lagged weather variable, $w_{t+1-j,s}$, standardised in the same manner as for the corresponding national scale analysis and summing over the specified number of months to form site specific values for the combined rainfall index (R_{ts}) and the combined temperature index (T_{ts}). Mathematically, the calculation for the temperature index with a maximum lag of 10 years can be expressed as:

$$T_{ts} = \sum_{j=1}^{120} b_j w_{t+1-j,s}.$$

Inclusion of the combined rainfall index and the combined temperature index as covariates in the the fine scale analysis implies that modelled populations in the North of Scotland will respond in a similar way to absolute temperature values as population in the South of England. In accordance with the view that we should focus on hypothesis H3 above, we have included an interaction between the combined temperature index and site-specific mean annual temperature in our models.

Model for fine-scale analysis

The population data at the site by year level are always counts of individuals in each year, hence it is appropriate to analyses these as generalised linear models (GLMs) with Poisson errors (to allow for the integer nature of the data) and a log link function (to ensure means values are always positive). To account for density dependence we will include the population in year -1 as a fixed effect in the models, and whilst variation between sites and years has been allowed for by the inclusion of categorical variables as random effects (making the models fitted generalised linear mixed models (GLMMs)).

References

Freeman S.N. & Newson S.E. (2008). On a log-linear approach to detecting ecological interactions in monitored populations. *Ibis*, **150**, 250-258.

Oliver, T., Roy, D. B., Hill, J. K., Brereton, T. and Thomas, C. D. (2010), Heterogeneous landscapes promote population stability. *Ecology Letters*, **13**, 473–484.

Perry M. & Hollis D. (2005) The generation of monthly gridded datasets for a range of climatic variables over the UK. *Int. J. Clim.* **25**, 1041–1054.

Roy D.B., Rothery P., Moss D., Pollard E. & Thomas J. A. (2001). Butterfly numbers and weather: predicting historical trends in abundance and the future effects of climate change. *Journal of Animal Ecology*, **70**, 201– 217.

Appendix 1 A list of the habitat types included in the fine scale analysis

Habitat class
Arable
Bare ground and quarries
Bracken
Broadleaved woodland
Coastal
Coniferous woodland
Fen
Grassland
Heath
Inland water
Montane
Urban/ suburban garden
